

Elastin-like polypeptide and γ -zein fusions significantly increase recombinant protein accumulation in soybean seeds

Jing Yang

Jilin Academy of Agricultural Sciences

Lu Niu

Jilin Academy of Agricultural Sciences

Hongli He

Jilin Normal University

Yunqing Cheng

Jilin Normal University

Xiaofang Zhong

Jilin Academy of Agricultural Sciences

Qianqian Zhao

Jilin Academy of Agricultural Sciences

Guojie Xing

Jilin Academy of Agricultural Sciences

Jianfeng Liu

Jilin Normal University

Xiangdong Yang (✉ xdyang020918@126.com)

Jilin Academy of Agricultural Sciences <https://orcid.org/0000-0002-7101-1626>

Research article

Keywords: Soybean seed, Fusion partner, Elastin-like polypeptide, γ -Zein, Recombinant protein

Posted Date: December 30th, 2019

DOI: <https://doi.org/10.21203/rs.2.19651/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Transgenic Research on May 8th, 2021. See the published version at <https://doi.org/10.1007/s11248-021-00258-7>.

Abstract

Background

Soybean seeds are an ideal host for the production of recombinant proteins because of their high content of proteins, long-term stability of seed proteins under ambient conditions, and easy establishment of efficient purification protocols.

Results

In this study, a polypeptide fusion strategy was applied to explore the capacity of elastin-like polypeptide (ELP) and γ -zein fusions in increasing the accumulation of the recombinant protein in soybean seeds. Transgenic soybean plants were generated to express the γ -zein- or ELP-fused green fluorescent protein (GFP) under the control of the soybean seed-specific promoter *BCSP*. Significant differences were observed in the accumulation of zein-GFP and GFP-ELP from that of the unfused GFP in transgenic soybean seeds based on the total soluble protein (TSP), despite the low-copy of T-DNA insertions and similar expression at the mRNA levels in selected transgenic lines. The average levels of zein-GFP and GFP-ELP accumulated in immature seeds of these transgenic lines were 0.99% and 0.29% TSP, respectively, compared with 0.07% TSP of the unfused GFP. In mature soybean seeds, the accumulation of zein-GFP and GFP-ELP proteins was 1.8% and 0.84% TSP, an increase of 3.91- and 1.82-fold, respectively, in comparison with that of the unfused GFP (0.46% TSP). Confocal laser scanning analysis showed that both zein-GFP and GFP-ELP were deposited in the protein storage vacuoles of transgenic seeds, while the unfused GFP was mainly distributed in the cytoplasm. Despite increased recombinant protein accumulation, there were no significant changes in the total protein and oil content in seeds between the transgenic and non-transformed plants, suggesting the possible presence of threshold limits of total protein accumulation in transgenic soybean seeds.

Conclusions

Overall, our results indicate that γ -zein and ELP fusions significantly increased the accumulation of the recombinant protein, but exhibited no significant influence on the total protein and oil content in soybean seeds.

1. Background

Soybean seeds are a practical expression system for the biotechnological production of recombinant proteins. Besides the general advantages of plant production systems, such as unlimited scale-up potentials, remarkably reduced production costs, eukaryotic post-translational modifications, and low contamination risks [1, 2], soybean seeds also possess several inherent benefits for recombinant protein expression [3]. For example, soybean seeds contain approximately 40% protein by dry mass (compared with 7%–12% in cereals such as maize, rice, and wheat), which provide ample and stable space for the deposition of recombinant proteins in seeds [3]. The general properties such as low protease activities

and water content in soybean seeds provide stable biochemical environments for long-term stable post-harvest storage of recombinant proteins without any detectable degradation [4-6]. Cunha et al. (2011a) reported that correct processing and stable accumulation of functional recombinant human coagulation factor IX (hFIX) could be maintained in soybean seeds for 6 years at room temperature [4]. Moreover, simple proteome, relative homogeneity, and less metabolites in soybean seeds also facilitate downstream processing and purification of recombinant proteins [2, 6]. To date, several recombinant pharmaceuticals and antigen vaccines, including *Escherichia coli* heat labile toxin B subunit (LTB) [7], anti-hypertensive peptide [8], human growth hormone [9], human coagulation factor IX [4], and microbicide against HIV cyanovirin-N [10], have been successfully expressed in soybean seeds. However, compared with that in other host plants such as rice seeds [11, 12], the accumulation of some recombinant proteins in soybean seeds is usually below the 1% total soluble protein (TSP) of economic threshold proposed by Kusnadi et al. (2015) [13], except one report in which the accumulation level of the recombinant protein reached 2.9% TSP in soybean seeds [9].

Multiple strategies have been developed and tested to increase the yield and stability of recombinant proteins in plant production systems. Among those, fusion protein technology showed great potential in enhancing recombinant protein accumulation and facilitating recovery and downstream purification [14-17]. Several fusion partners such as oil body-targeted oleosin, elastin-like polypeptide (ELP), 27-kDa maize γ -zein, and fungal hydrophobin I (HFBI) have been utilized to improve both the accumulation and stability of recombinant proteins [14, 15, 17-20]. γ -Zein, a member of the major prolamin storage family in maize seeds, contains a signal peptide with a "CGC" motif, proline-rich amphipathic region with eight "PPPVHL" repeats, and third domain with four additional cysteine residues at its N-terminal region. γ -Zein possesses an intrinsic protein body (PB)-forming property in the absence of tissue-specific factors [16]. When fused to a recombinant protein of interest, γ -zein can successfully trigger PB formation and induce high accumulation of recombinant proteins [16, 20-22]. For example, fusion of the human immunodeficiency virus negative factor (Nef) sequence to the N-terminal domain of γ -zein led to the accumulation of 1.5% TSP in transgenic tobacco leaves [23]. Elastin-like polypeptide is repeats of the amino acid "VPGXG" that can undergo a temperature-dependent reversible phase transition. The use of ELP as a fusion partner has also been reported to be capable of inducing PB formation in the leaves when targeted to the endoplasmic reticulum (ER) in transient expression assays [24-26]. The addition of ELP to the C-terminus of *Influenza* hemagglutinin resulted in its increased accumulation in both tobacco leaves and seeds compared with that of unfused hemagglutinin [15]. The expression of single-chain variable fragments (scFvs) in transgenic tobacco seeds even increased to 40-fold, with levels approaching 25% TSP, when fused with ELP at the C terminal [27].

Seed vacuoles are intracellular endpoints of the plant secretory pathways for proteolysis or storage of proteins. Different from the lytic vacuoles, as seed-specific organelles, the protein storage vacuoles (PSVs) are specialized for the deposition of large amounts of endogenous proteins [28-30]. The abundance of PSVs in the seeds makes it the preferred site for recombinant protein accumulation. PSV signal peptides have been used to direct recombinant proteins into PSVs of tobacco or soybean seeds in previous studies [8, 16, 31, 32]. Trafficking and storage of proteins to PSVs may depend on plant species,

tissues, and developmental stages [33]. As for PBs, although they may remain in the cytosols after formation and budding from ER, they can also be sequestered in the PSVs probably via an autophagy-like process for long-term storage, as showed in wheat, oats, barley, and tobacco [33].

In this study, we explored the capacity of two fusion partners ELP and γ -zein in increasing recombinant protein accumulation in soybean seeds, which naturally accumulate higher amounts of endogenous proteins than that of most other crops. In addition, the targeted deposition of the recombinant protein and the influence of recombinant protein accumulation on the total protein and oil content in transgenic soybean seeds were also analyzed. Our aim was to provide a practical and efficient protocol for the production of important pharmaceutical proteins or vaccine candidates with soybean seeds as an expression host.

2. Results

Molecular characterization of transgenic soybean plants

To evaluate the influence of the polypeptide fusion partners ELP and γ -zein on the accumulation of GFP in soybean seeds, three plant expression vectors harboring *GFP-ELP*, *zein-GFP*, and *GFP* were constructed and used for *Agrobacterium*-mediated genetic transformation (Fig. 1A). The soybean seed-specific promoter *BCSP*, which is specifically expressed in the cotyledon and embryonic axis of soybean seeds [34], was used to drive the expression of the *GFP* fusion genes or unfused *GFP*. A series of transformation experiments resulted in 15 zein-GFP, 18 GFP-ELP, and 23 GFP lines as confirmed using the LibertyLink strip and PCR analysis (Fig. S1A-D). In this study, all the transgenic plants grew normally and showed no differences in morphology compared with that of non-transformed plants (data not showed). Transgenic plants in the subsequent generations (T_1 - T_3) were further screened for their herbicide tolerance and by PCR analysis until homozygous transgenic lines were obtained. To ensure consistency in comparing accumulation levels of the recombinant proteins, we first analyzed the relative expression of *zein-GFP*, *GFP-ELP*, and *GFP* in the immature seeds at 45 days after flowering by quantitative RT-PCR. The results revealed extensive variability in the transcript levels across the independent transgenic soybean events (Fig. 1B), which might be caused by positional effects of transgene insertion within the genome, transgene copy number, and/or silencing mechanisms suppressing transgene expression [15]. However, in some transgenic events including DG-12, DG-44, DG49, DE-1, DE-9, DE-28, DF-4, DF-8, and DF-43, similar expression levels of the foreign genes at the mRNA levels were also observed at the same time point during seed development (Fig. 1B), likely due to the same *BCSP* promoter used in the study. Integration and copy number of the transgenes were further determined by Southern blotting. Hybridization signals were observed in the selected transgenic plants, while no positive signal was detected in the non-transformed plants, confirming the insertion of the foreign genes in the soybean genome (Fig. 2A). In these transgenic events, the copy number ranged from 1 to 3. To avoid influence of copy number and differential expression at the mRNA level of the transgenes on recombinant protein accumulation in different transgenic events, only the transgenic plants with a low-copy (single or double) of insertion and similar expression of the transgenes were selected for further investigation.

Expression and quantification of the recombinant GFP protein

Western blotting was conducted to analyze the expression of zein-GFP, GFP-ELP, and unfused GFP with the monoclonal anti-GFP antibodies. The band with an apparent molecular weight of 26.8 kDa representing GFP was observed in the five GFP transgenic plants tested (Fig. 2B). Comparatively, the zein-GFP and GFP-ELP fusion proteins were detected as soluble forms of approximately 51.8 and 40.5 kDa, respectively, which represented the monomer of each recombinant protein of the expected size (Fig. 2B). However, a 26.8-kDa band was also detected in both zein-GFP and GFP-ELP seed samples, probably due to proteolytic cleavage between the GFP and its fusion partners (Fig. 2B). These results indicated that the GFP antibody can effectively bind to the γ -zein- or ELP-fused GFP protein in proper configurations and confirmed the translation of both recombinant proteins in the transgenic soybean plants. Recombinant protein accumulation in soybean seeds was further quantified using ELISA. As shown in Fig. 3A, significant differences were observed in the accumulation of zein-GFP or GFP-ELP and unfused GFP in immature soybean seeds. All three zein-GFP lines DG-12, DG-44, and DG49 displayed significantly higher protein accumulation with an average of 0.99% TSP (0.84%–1.2%), than that of both GFP-ELP and GFP proteins, which showed 0.29% TSP (0.25%–0.34%) and 0.07% TSP (0.06%–0.08%), respectively (Fig. 3A). Although lower than that of zein-GFP in immature seeds, the average expression of GFP-ELP was still considerably higher than that of the unfused GFP. Consistent results were also observed in mature seeds (Fig. 3B). In this case, the average accumulation of zein-GFP and GFP-ELP proteins reached 1.8% (1.42%–2.08%) and 0.84% TSP (0.79%–0.85%), an increase of 3.91- and 1.82-fold, respectively, compared with that of the unfused GFP (0.46% TSP) (Fig. 3B). Furthermore, zein-GFP also showed higher accumulation than GFP-ELP in mature seeds. Overall, these observations demonstrated that γ -zein and ELP polypeptide fusions significantly enhance the expression and accumulation of GFP in soybean seeds, and γ -zein appears to have a greater influence on GFP accumulation than ELP under our experiment conditions.

Subcellular targeting of GFP with different fusion partners in soybean seeds

Confocal laser scanning analysis was performed to analyze subcellular targeting of the GFP fused with γ -zein or ELP in immature soybean seeds. Compared with that in the untransformed plants, green fluorescence signal was observed in all transgenic lines, which was attributed to the expression of GFP. In zein-GFP- and GFP-ELP-expressing plants, abundant small spherical particles representing PSVs were observed in immature soybean seeds (Fig. 4), while there were, if any, a few fluorescence signals in the same cellular compartments of the unfused GFP-expressing seeds. Previous studies have showed that both γ -zein and ER-targeting ELP fusion could induce the formation of PBs in vegetative plant tissues [16, 20-22, 24-26]. However, the PBs originating from the endomembrane system can also be deposited in the PSVs via an autophagy-like process [33], depending on plant tissue and developmental stage. In our study, γ -zein and ELP-induced PBs were sequestered into PSVs of soybean seeds despite the absence of PSV localization signal in the fusion proteins.

Influence of recombinant protein accumulation on soybean seed protein and oil content

Soybean seeds are characteristically high in proteins. To evaluate the effects of accumulation of the recombinant proteins on the total protein and oil content in soybean seeds, we analyzed the changes in the content of total protein and oil in transgenic soybean seeds expressing zein-GFP, GFP-ELP, and GFP proteins. The results showed that there were no significant differences in seed protein and oil content between the transgenic and non-transformed soybean seeds (Fig. 5A, B). Increased recombinant protein accumulation does not appear to influence the total seed protein (native and recombinant) content, implying that there appears to be a threshold limit on the total level of soybean protein, due to the competition for resources in protein biosynthesis, trafficking, and deposition between the recombinant and native proteins. We suppose that increased recombinant protein accumulation in soybean seeds is likely accompanied by the displacement of native storage proteins. The influence of recombinant protein accumulation on total protein and oil content needs to be further investigated in a larger field experiment.

3. Discussion

There are two major limitations to the potential production system of recombinant proteins: low accumulation levels and lack of efficient purification [17]. Sufficient accumulation of recombinant proteins in host plant is a major requirement for their commercial-scale production. Currently, polypeptide fusion strategy is an efficient way to enhance recombinant protein yield and also provides a simple means for the purification of recombinant proteins [14-17]. In this study, two polypeptide partners γ -zein and ELP were evaluated for their capability to enhance recombinant GFP accumulation in soybean seeds. To avoid the influence of transgene copy number, which might have a positive or negative association with expression level, only the transgenic lines with low-copy insertion (single or double) were utilized in the expression analysis. Moreover, the transcription levels of the transgenes were determined before quantification of the recombinant proteins. The results revealed that the accumulation of zein-GFP and GFP-ELP was significantly increased compared with that of unfused GFP in both immature and mature soybean seeds. These results are consistent with those of previous studies, which reported that the expression of γ -zein or ELP-fused recombinant proteins was higher than that of proteins without fusion partners in plant leaves or seeds [15, 16, 23, 24, 26, 27]. Enhanced accumulation of the recombinant proteins might result from increased transportation and deposition in cell compartments such as PBs or PSVs, which can protect recombinant proteins from proteolytic digestion. Moreover, accumulation enhancement by the polypeptide fusion is also dependent on the specific fusion partner. In contrast to ELP, γ -zein appears to have a greater influence on GFP accumulation in both immature and mature transgenic soybean seeds. A similar observation was reported by Phan et al. (2014) [17], in which case, ELP increased the accumulation of GFP to a less extent than HFBI [15].

Accumulation of the recombinant proteins is regulated by several steps at the transcriptional (promoters and transcription factors), translational, and post-translational levels (modification, processing trafficking, and deposition). At the endpoint of the plant secretory pathway, PSVs present the major site for the accumulation and long-term storage of protein in the seeds [15]. The temporal abundance of PSVs is also important for the accumulation of high-level endogenous proteins in soybean seeds. Soybean seeds are capable of producing a different class of functional recombinant proteins correctly targeted to

the PSVs at appropriate levels [30]. In this study, PSVs were observed in the seeds of transgenic plants expressing zein-GFP and GFP-ELP without a PSV localization signal. The PBs triggered by γ -zein or ELP can be sequestered in the PSVs independent of the PSV-sorting determinants, via an autophagy-like process for long-term storage [33], especially in dicots such as soybean seeds in which endogenous storage proteins are mainly deposited in PSVs. Compartmentalization of the recombinant protein in PBs or PSVs not only avoids proteolysis, but also increases the yield and stability of recombinant proteins [4].

In this study, we also evaluated the effects of recombinant protein accumulation on the total seed protein content in the host plant. This is especially important for soybean seeds, which possess a high level of endogenous storage protein. Our results showed that recombinant protein accumulation has no detectable effect on total protein storage, suggesting that an intrinsic threshold mechanism exists to control total protein accumulation. Competition in biosynthesis, transportation, and deposition between the recombinant and native proteins for resources may affect the deposition of amount of foreign proteins in soybean seed. Previous studies have showed that the suppression of the α and α' subunits of β -conglycinin does not influence the total oil and protein content and ratio of soybean seeds, and the decrease in β -conglycinin protein can be compensated by increased accumulation of glycinin [35]. The expression and accumulation of recombinant proteins might simultaneously result in decreased native protein storage in soybean seeds. This observation provides a cue that it might further increase the accumulation of desired recombinant proteins in soybean seeds by suppressing native storage protein via RNA interference or gene editing. Overall, our results demonstrated the practicability of ELP and γ -zein as fusion partners to enhance the accumulation of recombinant protein and exhibit no significant influence on the total protein and oil content in soybean seeds.

4. Conclusions

In this study, a polypeptide fusion strategy was applied to explore the capacity of elastin-like polypeptide (ELP) and γ -zein fusions in increasing the accumulation of the recombinant protein in soybean seeds. And the recombinant protein was accumulated in transgenic soybean seeds. Both zein-GFP and GFP-ELP were deposited in the protein storage vacuoles of transgenic seeds, while the unfused GFP was mainly distributed in the cytoplasm. There were no significant changes in the total protein and oil content in seeds between the transgenic and non-transformed plants.

5. Methods

Vector construction and *Agrobacterium*-mediated transformation of soybean

All cloning steps were carried out using the binary vector pTF101. The 1413-bp seed-specific promoter of β -conglycinin alpha subunit (*BCSP*) was amplified from soybean as previously described [36] and sub-cloned into the *Hind*III/*Xba*I sites of the compatible pTF101 plasmid. The genes encoding maize 27 kDa γ -zein (Genbank No. AF371261.1) and 13.56 kDa ELP proteins were synthesized and optimized based on soybean codon usage (GenScript USA Inc., NJ, USA) to remove cryptic splice sites and rare codons, and

then inserted into the binary vector under the control of the promoter *BCSP*. The coding sequence of the green fluorescent protein (GFP) was then cloned upstream of ELP or downstream of γ -zein sequences to generate the transformation vectors pTF101-GFP-ELP and pTF101-zein-GFP. In construct pTF101-GFP-ELP, the ER localization signal KDEL sequence was added at the C-terminus of ELP, while in construct pTF101-zein-GFP, the enterokinase cleavage site (DDDDK) was located between γ -zein and GFP instead (Fig. 1A). For the control vector pTF101-GFP, the *GFP* gene was cloned into the *Xba*I/*Sac*I sites between *BCSP* and nopaline synthase (*NOS*) terminator. All three vector constructs contained a phosphinothricin acetyltransferase selectable marker gene (*bar*) for the selection of glufosinate-resistant transgenic events. The soybean genotype P03, provided by Prof. Bao Peng from the Jilin Academy of Agricultural Sciences, China, was used for *Agrobacterium*-mediated transformation as previously described by Yang et al. (2017) [37]. The primary transformants were transferred into a greenhouse for subsequent molecular screening.

Molecular analysis of the transgenic plants

The T_0 transgenic plants were first screened using the LibertyLink[®] strip according to the instructions of the manufacturer (EnviroLogix Inc., ME, USA). Polymerase chain reaction (PCR) was performed to confirm the presence of the transgenes using the primers BG-F1/BG-R1 (5'-CTCACCATCGCTCAACACATT TC-3', 5'-TGTCTTGTAGTTCCCGTCGTCC-3'), which anneal with *GFP* and its upstream *BCSP* sequences, respectively. For the subsequent generations, the transgenic plants were further screened by both glufosinate (1500 mg/L) spraying and PCR analysis. Southern hybridization was conducted to determine transgene copy numbers in soybean plants. Total genomic DNA was extracted from the young leaves of T_3 transgenic plants with a modified procedure using cetyltrimethylammonium bromide at a high salt concentration as described previously [38]. Approximately 30 μ g of genomic DNA was digested overnight with *Xba*I, separated on 1% agarose gel, and transferred onto a positively charged nylon membrane (GE Amersham, USA) according to the standard protocols. The 714-bp *GFP* probe was obtained by PCR amplification using the primers GFP-F1/GFP-R1 (5'-ATGAGTAAAGGAGAAGAAGACTTTTCA-3', 5'-TTTGTATAGTTCATCCATGC

CATGT-3') and labeled using digoxigenin (DIG)-high prime (Roche, Germany). Hybridization, membrane washing, and subsequent chemical staining with BCIP/NBT were conducted as described previously [37].

Quantitative reverse transcription (qRT)-PCR

Forty-five days after flowering, immature transgenic soybean seeds were sampled to determine the relative expression of the *GFP* transcripts by qRT-PCR. Total RNA was extracted from the immature seeds (\approx 10 mm in length) of T_3 transgenic plants using the EasyPure Plant RNA Kit (TransGen Biotech, Beijing, China), according to the manufacturer's protocol. Genomic DNA removal and first-strand cDNA synthesis were carried out using TransScript[®] One-Step gDNA Removal and cDNA Synthesis SuperMix (TansGen Biotech). Amplifications were performed on ABI 7900HT (Applied Biosystems, USA) using TransStart[®] Top Green qPCR SuperMix (TansGen Biotech). The reaction conditions were as follows: 50°C for 2 min,

95°C for 10 min, and 35 cycles of 95°C for 2 min, 60°C for 30 s, and 72°C for 30 s. The specific primers GFP-F2/GFP-R2 (5'-TACGTGCAGGAGAGGACCA T-3', 5'-ACTTGTGGCCGAGGATGTTT-3') were designed based on the *GFP* sequence. The constitutively expressed native soybean gene *GmACT6* (GenBank No. NM_001289231) was amplified with the primers GmACT-F1/GmACT-R1 (5'-ATCTTGACTGAGCGTGGTTATTCC-3', 5'-GCTGGTCTGGCTGTCTCC-3'). The expression levels were calculated using the relative quantification ($2^{-\Delta Ct}$) method and the data were normalized using the expression level of the internal control *GmACT6*. All experiments were conducted with three biological replicates.

Western blotting and GFP quantification analysis

The T₃ transgenic lines carrying a low-copy (single or double) transgene insertion and with similar mRNA expression of the transgenes in the immature seeds at 45 days after flowering were used for western blotting and enzyme-linked immunosorbent assay (ELISA). Immature seeds of similar size (≈10 mm in length) were harvested from the transgenic plants expressing zein-GFP, GFP-ELP, or unfused GFP. Total soluble protein was extracted with extraction buffer (100 mM NaCl, 10 mM EDTA, 200 mM Tris-HCl, 0.05% Tween-20, 0.1% SDS, 14 mM β-mercaptoethanol, 400 mM sucrose, and 2 mM phenylmethane sulfonyl fluoride). Protein quantification was performed using the Bradford method [39]. The protein samples were then separated by 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis and electrotransferred on to a PVDF membrane (GE Healthcare, USA). After blocking with 5% (w/v) fat-free milk powder, the membranes were blotted with monoclonal anti-GFP antibodies (1:500 dilution, Abcam, UK) for 2 h at room temperature. Antibody binding was then detected by the addition of horseradish peroxidase (HRP)-conjugated goat-anti-rabbit IgG (1:5000 dilution, Abcam). After extensive washing, the signal was visualized using the Biodlight™ Western Chemiluminescent HRP substrate (Bioworld Technology, Inc. USA). To quantify GFP, the protein samples from immature or mature soybean seeds were subjected to ELISA with purified GFP (Abcam) as the standard protein. The wells of the ELISA plates were coated with diluted protein extracts and standard protein in the coating buffer (15 mM Na₂CO₃, 35 mM NaHCO₃, and 3 mM NaN₃; pH 9.6). After blocking with 5% dried skimmed milk in phosphate buffer containing 0.1% Tween 20, the monoclonal anti-GFP antibodies (1:500 dilution) were added into the wells and incubated for 1 h at room temperature. The HRP-conjugated secondary antibodies were then added at a dilution of 1:5000 and incubated for 1 h. Optical density (OD) of each reaction mixture was measured at 450 nm (ELx800; BioTek, USA) with tetramethylbenzidine (TMB) as the substrate.

Confocal laser scanning

Immature seeds were collected at 45 days after flowering and tissue sections were prepared by hand. The samples were immediately imaged using a Leica TCS SP2 confocal laser scanning inverted microscope (Leica Microsystems, Germany) equipped with a 63× water immersion objective. To visualize GFP fluorescence, excitation with 488-nm argon laser was performed and the emission was detected at 500–530 nm. The collected images were analyzed using Leica Application Suite for Advanced Fluorescence (LAS AF, V2.3.5; Leica Microsystems).

Protein and oil content analyses in transgenic seeds

Protein and oil content in the transgenic and non-transformed soybean seeds were analyzed as described previously [37]. Briefly, mature soybean seeds (200 mg) were ground in a grinder after drying at 60°C. The level of total nitrogen was determined using LECO CHN 2000 analyzer (LECO, St. Joseph, MI) as described by Hwang et al. (2014) [40]. The seed protein content was calculated based on the dry weight basis. The seed oil content was determined by pulsed NMR (Maran Pulsed NMR, Resonance Instruments, Oxfordshire, UK), followed by the field induction decay-spin echo procedure [41].

Statistical analysis

All statistical analyses were performed using SPSS software v.17.0 (SPSS Inc., Chicago, IL, USA). Differences in the average values of samples were tested using the analysis of variance (ANOVA).

Abbreviations

ELP: elastin-like polypeptide; GFP: green fluorescent protein; TSP: total soluble protein; hFIX: human coagulation factor IX; LTB: labile toxin B; HIV: Human Immunodeficiency Virus; HFBI: fungal hydrophobin I; PB: protein body; ER: endoplasmic reticulum; PSV: sprout storage vacuoles; *BCSP*: β -conglycinin alpha subunit; *NOS*: nopaline synthase; PCR: Polymerase chain reaction; DIG: digoxigenin; qRT: Quantitative reverse transcription; EDTA: Ethylene Diamine Tetraacetic Acid; HRP: horseradish peroxidase; OD: Optical density; TMB: tetramethylbenzidine; PVDF: Polyvinylidene Fluoride

Declarations

Acknowledgments

We thank that Prof. Bao Peng from the Jilin Academy of Agricultural Sciences, China, kindly provided the soybean cultivar P03. We also would like to thank Editage (www.editage.cn) for English language editing.

Author Contribution

XDY and JFL designed the experiments. JY and LN performed the experiments and drafted the manuscript. HLH, QQZ, and GJX conducted the *Agrobacterium*-mediated transformation experiments. YQC conducted the confocal laser scanning analysis. XFZ analysed the experimental data. All authors have read and approved the final version of the manuscript.

Funding

This study was financially supported by the National Natural Science Foundation of China (31671764) and the China National Novel Transgenic Organisms Breeding Project (2016ZX08004-003). These funds played an important role in experimental reagents, consumables, and also provided strong support for preparing manuscript.

Availability of data and materials

All the datasets used and/or analysed throughout this study can be available from the corresponding authors on reasonable request.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors claim no conflict of interest regarding publication of this paper.

References

1. Kawakatsu T, Takaiwa F. Cereal seed storage protein synthesis: fundamental processes for recombinant protein production in cereal grains. *Plant Biotechnology Journal*. 2013;8:939-953.
2. Lau OS, Sun SS. Plant seeds as bioreactors for recombinant protein production. *Biotechnology Advances*. 2009;27:1015-1022.
3. Hudson LC, Garg R, Bost KL, Piller KJ. Soybean seeds: a practical host for the production of functional subunit vaccines. *Biomed Res Int*. 2014;340804.
4. Müntz K. Deposition of storage proteins. *Plant Molecular Biology*. 1998;38:77-99.
5. Morandini F, Avesani L, Bortesi L, Van DB, De WK, Arcalis E, Bazzoni F, Santi L, Brozzetti A, Falorni A. Non-food/feed seeds as biofactories for the high-yield production of recombinant pharmaceuticals. *Plant Biotechnology Journal*. 2011;9:911-921.
6. Moravec T, Schmidt MA, Herman EM, Woodfordthomas T. Production of *Escherichia coli* heat labile toxin (LT) B subunit in soybean seed and analysis of its immunogenicity as an oral vaccine. *Vaccine*. 2007;25:1647-1657.
7. Yamada Y, Nishizawa K, Yokoo M, Zhao H, Onishi K, Teraishi M, Utsumi S, Ishimoto M, Yoshikawa M. Anti-hypertensive activity of genetically modified soybean seeds accumulating novokin. *Peptides*. 2008;29:331-337.
8. Cunha NB, Murad AM, Cipriano TM, Araújo ACG, Aragão FJL, Leite A, Vianna GR, McPhee TR, Souza GHMF, Waters MJ. Expression of functional recombinant human growth hormone in transgenic soybean seeds. *Transgenic Research*. 2011b;20:811-826.
9. O'Keefe BR, Murad AM, Vianna GR, Ramessar K, Saucedo CJ, Wilson J, Buckheit KW, Cunha NB, Araújo ACG, Lacorte CC. Engineering soya bean seeds as a scalable platform to produce cyanovirin-N, a non-ARV microbicide against HIV. *Plant Biotechnology Journal*. 2015;13:884-892.

10. He Y, Ning T, Xie T, Qiu Q, Zhang L, Sun Y, Jiang D, Fu K, Yin F, Zhang W. Large-scale production of functional human serum albumin from transgenic rice seeds. *Proc Natl Acad Sci USA*. 2011;108(47):19078-19083.
11. Zhang D, Lee HF, Pettit SC, Zaro JL, Ning H, Shen WC. Characterization of transferrin receptor-mediated endocytosis and cellular iron delivery of recombinant human serum transferrin from rice (*Oryza sativa* L.). *BMC Biotechnology*. 2012;12:92.
12. Kusnadi AR, Nikolov ZL, Howard JA. Production of recombinant proteins in transgenic plants: practical considerations. *Biotechnology & Bioengineering*. 2015;56:473-484.
13. Conley AJ, Joensuu JJ, Richman A, Menassa R. Protein body-inducing fusions for high-level production and purification of recombinant proteins in plants. *Plant Biotechnology Journal*. 2011;9:419-433.
14. Gutiérrez SP, Saberianfar R, Kohalmi SE, Menassa R. Protein body formation in stable transgenic tobacco expressing elastin-like polypeptide and hydrophobin fusion proteins. *BMC Biotechnology*. 2013; 13:40.
15. Hofbauer A, Peters J, Arcalis E, Rademacher T, Lampel J, Eudes F, Vitale A, Stoger E. The induction of recombinant protein bodies in different subcellular compartments reveals a cryptic plastid-targeting signal in the 27-kDa γ -zein sequence. *Frontiers in Bioengineering and Biotechnology*. 2014;2:67.
16. Phan HT, Hause B, Hause G, Arcalis E, Stoger E, Maresch D, Altmann F, Joensuu J, Conrad U. Influence of elastin-like polypeptide and hydrophobin on recombinant hemagglutinin accumulations in transgenic tobacco plants. *Plos One*. 2014;9:e99347.
17. Floss DM, Schallau K, Rosejohn S, Conrad U, Scheller J. Elastin-like polypeptides revolutionize recombinant protein expression and their biomedical application. *Trends in Biotechnology*. 2010;28:37-45.
18. Joensuu JJ, Conley AJ, Lienemann M, Brandle JE, Linder MB, Menassa R. Hydrophobin fusions for high-level transient protein expression and purification in *Nicotiana benthamiana*. *Plant Physiology*. 2010;152:622-633.
19. Torrent M, Llompart B, Lasserre-Ramassamy S, Llop-Tous I, Bastida M, Marzabal P, Westerholm-Parvinen A, Saloheimo M, Heifetz PB, Ludevid MD. Eukaryotic protein production in designed storage organelles. *BMC Biology*. 2009;7:5.
20. Mainieri D, Rossi M, Archinti M, Bellucci M, Marchis FD, Vavassori S, Pompa A, Arcioni S, Vitale A. Zeolin. A new recombinant storage protein constructed using maize γ -zein and bean phaseolin. *Plant Physiology*. 2004;136:3447-3456.
21. Joseph M, Ludevid MD, Torrent M, Rofidal V, Tauzin M, Rossignol M, Peltier JB. Proteomic characterisation of endoplasmic reticulum-derived protein bodies in tobacco leaves. *BMC Plant Biology*. 2012;12:36.
22. Virgilio MD, Marchis FD, Bellucci M, Mainieri D, Rossi M, Benvenuto E, Arcioni S, Vitale A. The human immunodeficiency virus antigen Nef forms protein bodies in leaves of transgenic tobacco when fused to zeolin. *Journal of Experimental Botany*. 2008;59(10): 2815-2829.

23. Ceresoli V, Mainieri D, Fabbro MD, Weinstein R, Pedrazzini E. A fusion between domains of the human bone morphogenetic protein-2 and maize 27 kD γ -zein accumulates to high levels in the endoplasmic reticulum without forming protein bodies in transgenic tobacco. *Frontiers in Plant Science*. 2016;7:358.
24. Reza S, Amirali S, Joensuu JJ, Kohalmi SE, Rima M. Protein bodies in leaves exchange contents through the endoplasmic reticulum. *Frontiers in Plant Science*. 2016;7: 693.
25. Saberianfar R, Menassa R. Protein bodies - How the ER deals with high accumulation of recombinant proteins. *Plant Biotechnology Journal*. 2017;15:671-673.
26. Scheller J, Leps M, Conrad U. Forcing single-chain variable fragment production in tobacco seeds by fusion to elastin-like polypeptides. *Plant Biotechnology Journal*. 2006;4:243-249.
27. Jolliffe NA, Craddock CP, Frigerio L. Pathways for protein transport to seed storage vacuoles: Figure 1. *Biochemical Society Transactions*. 2005;33(5):1016-1018.
28. Boothe J, Nykiforuk C, Shen Y, Zaplachinski S, Szarka S, Kuhlman P, Murray E, Morck D, Moloney MM. Seed-based expression systems for plant molecular farming. *Plant Biotechnology Journal*. 2010;8: 588-606.
29. Vianna GR, Cunha NB, Murad AM, Rech EL. Soybeans as bioreactors for biopharmaceuticals and industrial proteins. *Genetics & Molecular Research Gmr*. 2011;10(3):1733-1752.
30. Leite A, Kemper EL, da Silva MJ, Luchessi AD, Siloto RMP, Bonaccorsi. Expression of correctly processed human growth hormone in seeds of transgenic tobacco plants. *Molecular Breeding*. 2000;6(1):47-53.
31. Kim WS, Krishnan HB. Expression of an 11-kDa methionine-rich delta-zein in transgenic soybean results in the formation of two types of novel protein bodies in transitional cells situated between the vascular tissue and storage parenchyma cells. *Plant Biotechnology Journal*. 2004;2(3):199-210.
32. Ibl V, Stoger E. The formation, function and fate of protein storage compartments in seeds. *Protoplasma*. 2012;249(2):379-392.
33. Imoto Y, Yamada T, Kitamura K, Kanazawa A (2008) Spatial and temporal control of transcription of the soybean beta-conglycinin alpha subunit gene is conferred by its proximal promoter region and accounts for the unequal distribution of the protein during embryogenesis. *Genes & Genetic Systems* 83:469-476
34. Kinney AJ, Herman JEM (2001) Cosuppression of the α subunits of β -conglycinin in transgenic soybean seeds induces the formation of endoplasmic reticulum-derived protein bodies. *The Plant Cell* 13(5):1165-1178
35. Yang J, Xing G, Niu L, He H, Guo D, Du Q, Qian X, Yao Y, Li H, Zhong X et al. Improved oil quality in transgenic soybean seeds by RNAi-mediated knockdown of *GmFAD2-1B*. *Transgenic Research*. 2018;27:155–166.
36. Yang X, Niu L, Zhang W, He H, Yang J, Xing G, Guo D, Du Q, Qian X, Yao Y, et al. Robust RNAi-mediated resistance to infection of seven potyvirids in soybean expressing an intron hairpin *Nlb* RNA. *Transgenic Research*. 2017;26:665-676.

37. Telzur N, Abbo S, Myslabodski D, Mizrahi Y. Modified CTAB procedure for DNA isolation from Epiphytic cacti of the *Genera Hylocereus* and *Selenicereus* (Cactaceae). *Plant Molecular Biology Reporter*. 1999;17:249-254.
38. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248–254
39. Hwang EY, Song QJ, Jia GF, Specht JE, Hyten DL, Costa J, Cregan PB (2014) A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics* 15:1. doi.org/10.1186/1471-2164-15-1
40. Rubel G (1994) Simultaneous determination of oil and water contents in different oil seeds by pulsed nuclear magnetic resonance. *Journal of American Oil Chemists' Society* 71:1057-1062

Figures

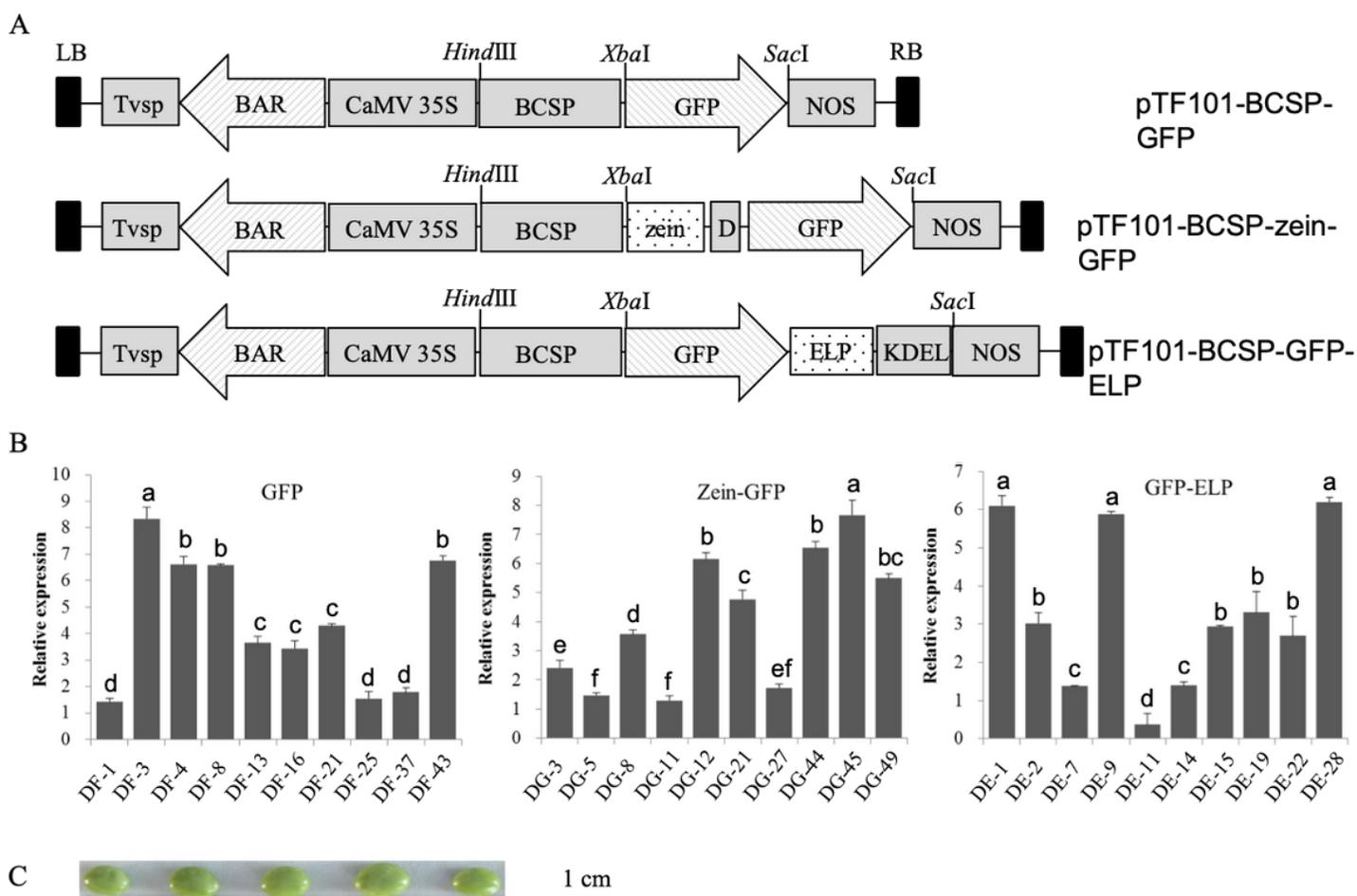


Figure 1

Plant expression constructs and generation of transgenic plants. (A) Schematic representation of the vector constructs used for *Agrobacterium*-mediated genetic transformation. BCSP, soybean seed-specific promoter of β -conglycinin alpha subunit; GFP, green fluorescent protein; Zein, maize 27 kDa γ -zein; ELP,

13.56 kDa elastin-like polypeptide; D, enterokinase cleavage site (DDDDK); KDEL, ER localization signal. The selectable marker BAR driven by the constitutive promoter CaMV 35S was used to select glufosinate-resistant transgenic plants. (B) Quantification of the GFP fusion genes or GFP alone at the mRNA level in immature seeds of representative transgenic soybean plants. RNA was extracted from the immature seed of T3 transgenic plants and GFP mRNA expression was determined using the $2^{-\Delta Ct}$ method after normalization to the expression of the internal control GmACTIN6. The data are presented as the mean of three biological replicates with error bars indicating the standard error. Letters indicate significant difference at $P < 0.05$. (C) Immature soybean seeds collected at 45 days after flowering for qRT-PCR and protein quantification analysis.

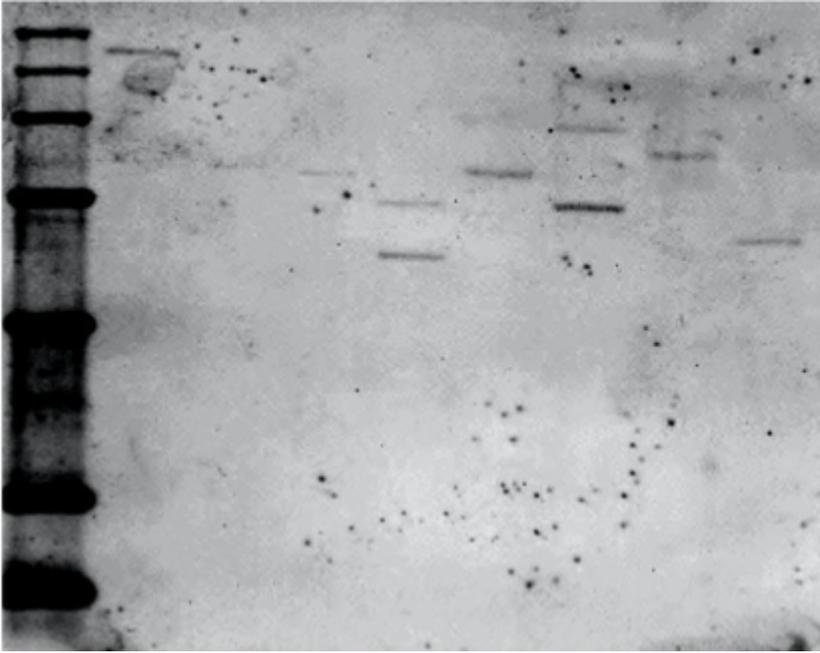


Figure 2

Integration and expression analysis of the transgenes. (A) Southern blotting of transgenic plants. Genomic DNA ($\times 30 \mu\text{g}$) was digested completely with XbaI and hybridized with a DIG-labeled GFP probe. M, DNA marker; Ctl+, positive control; Nt-P03, non-transformed plants; DF-1, DF-3, DF-4, DF-8, DF-21, DF-43, T3 GFP-expressing plants; DG-3, DG-8, DG-12, DG-44, DG-45, DG-49, T3 zein-GFP-expressing plants; DE-1, DE-2, DE-9, DE-14, DE-15, DE-28, T3 GFP-ELP-expressing plants. (B) Western blotting analysis showed the presence and accumulation of the GFP protein in immature transgenic seeds. Non-transformed Nt-P03 soybean plants as the negative control and purified GFP (Abcam, UK) as the positive control.

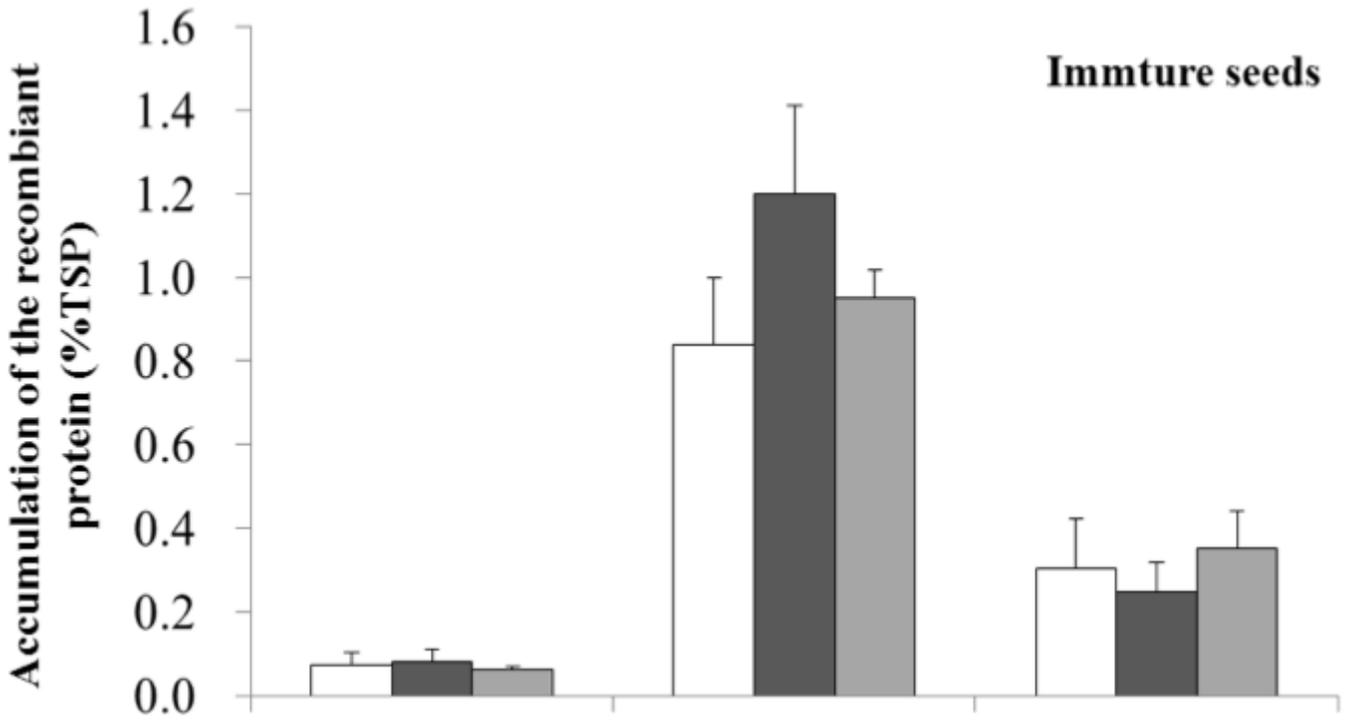


Figure 3

Quantification of the recombinant GFP protein in immature (A) and mature (B) soybean seeds. The total protein was extracted from zein-GFP-, GFP-ELP-, or GFP-expressing seeds at 45 days after flowering or mature and dry soybean seeds. GFP accumulation in immature soybean seeds was analyzed using the ELISA. Each column represents the mean of three independent biological replicates with the error bars indicating the standard error. The asterisks indicate significant differences at $P < 0.01$.

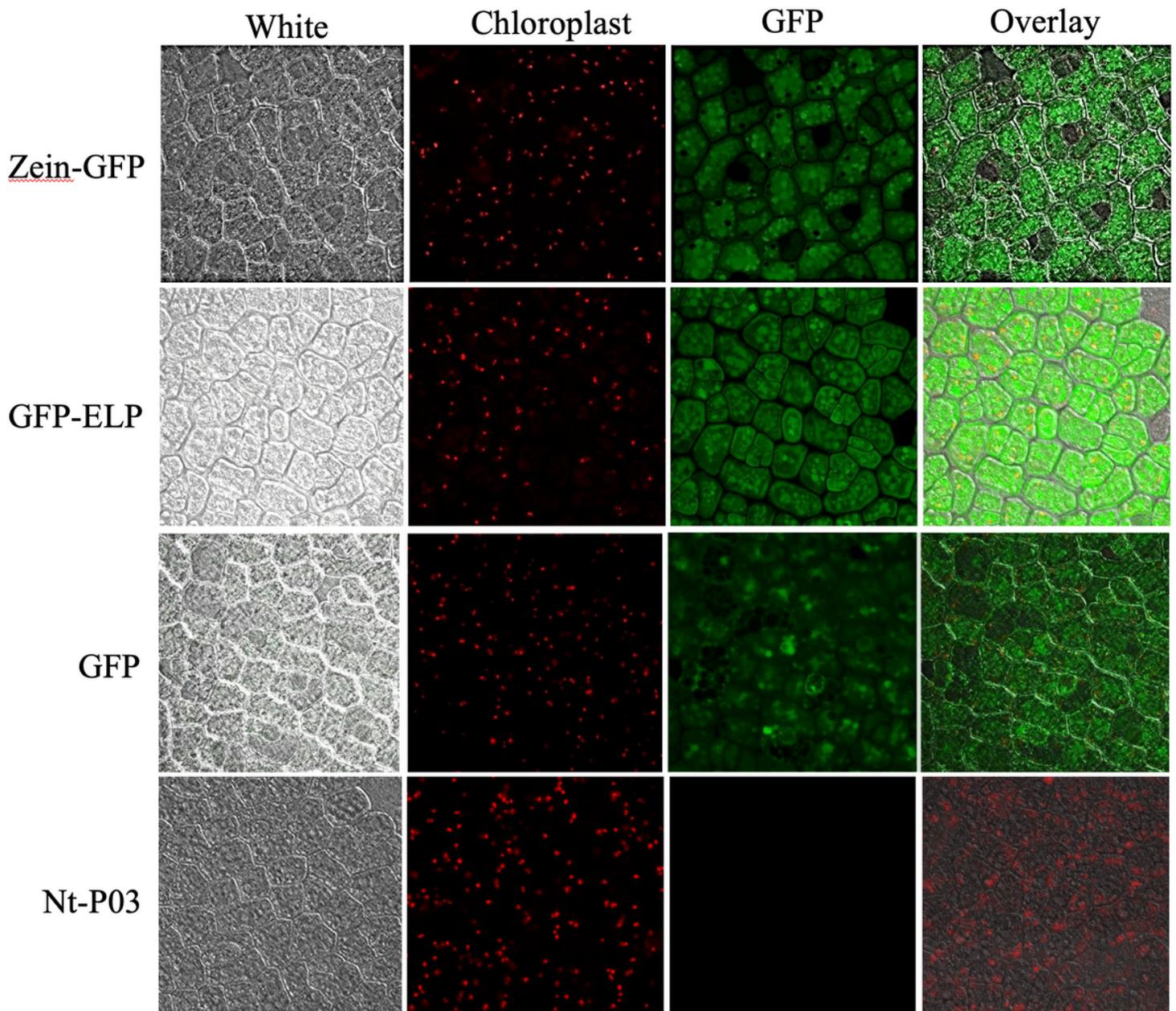


Figure 4

Laser confocal analysis of zein- or ELP-fused and unfused GFP in transgenic soybean seeds. Tissue samples were collected from immature soybean seeds at 45 days after flowering. Sample sections were imaged using the Leica TCS SP2 confocal laser scanning inverted microscope with a 63× water immersion objective (excitation wavelength = 488 nm, emission wavelength = 500–530 nm).

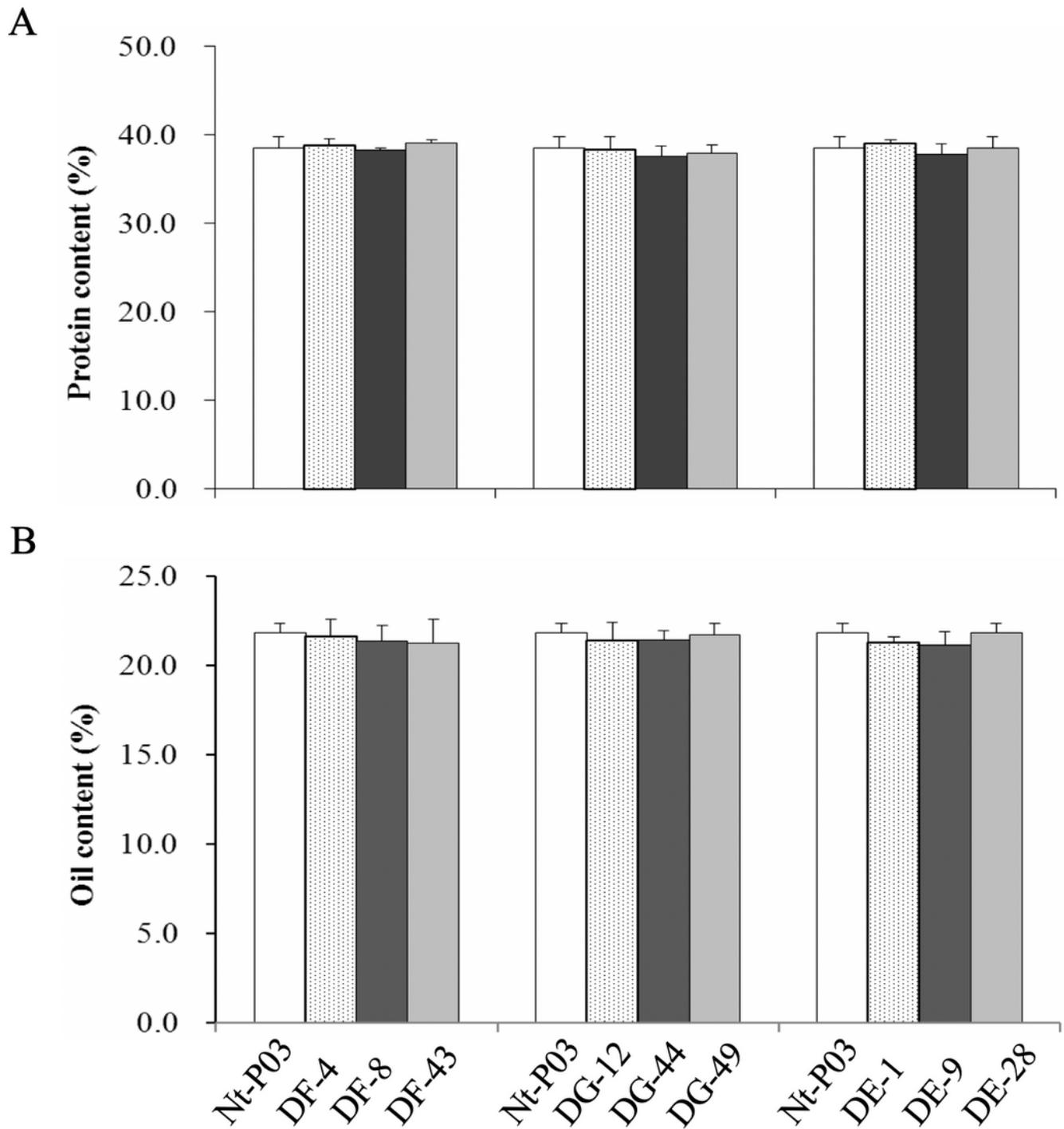


Figure 5

Influence of recombinant protein accumulation on the total protein and oil content in mature soybean seeds. The mature T3 soybean seeds were harvested and total protein and oil content was analyzed based on the dry weight. The data are presented as the mean of three biological replicates.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile.docx](#)